"Worth a thousand words"

Hawk, S.T.; Kleef, G.A. van; Fischer, A.H.; Schalk, J.

Published in:
Emotion

DOI:
10.1037/a0015178

Emotion
2009, Vol. 9, No. 3, 293–305

1528-3542/09/$12.00 DOI: 10.1037/a0015178

"Worth a Thousand Words": Absolute and Relative Decoding of Nonlinguistic Affect Vocalizations

Skyler T. Hawk, Gerben A. van Kleef, Agneta H. Fischer, and Job van der Schalk
University of Amsterdam

The authors compared the accuracy of emotion decoding for nonlinguistic affect vocalizations, speech-embedded vocal prosody, and facial cues representing 9 different emotions. Participants ($N = 121$) decoded 80 stimuli from 1 of the 3 channels. Accuracy scores for nonlinguistic affect vocalizations and facial expressions were generally equivalent, and both were higher than scores for speech-embedded prosody. In particular, affect vocalizations showed superior decoding over the speech stimuli for anger, contempt, disgust, fear, joy, and sadness. Further, specific emotions that were decoded relatively poorly through speech-embedded prosody were more accurately identified through affect vocalizations, suggesting that emotions that are difficult to communicate in running speech can still be expressed vocally through other means. Affect vocalizations also showed superior decoding over faces for anger, contempt, disgust, fear, sadness, and surprise. Facial expressions showed superior decoding scores over both types of vocal stimuli for joy, pride, embarrassment, and "neutral" portrayals. Results are discussed in terms of the social functions served by various forms of nonverbal emotion cues and the communicative advantages of expressing emotions through particular channels.

Keywords: emotion decoding, nonlinguistic affect vocalizations, vocal prosody, facial expressions, social functions of emotion

Supplemental materials: http://dx.doi.org/10.1037/a0015178.supp

Nonverbal signals are important for coordinating social life, because people often interpret such behaviors as symptomatic of expressers' emotions, attitudes, and/or behavioral intentions. Theorists have proposed emotion-related cues to fulfill several social functions that regulate interpersonal interactions (e.g., Fischer & Manstead, 2008; Fridlund, 1994; Keltner & Haidt, 1999; Scherer, 1980, 1988, 1994; Van Kleef, De Dreu, & Manstead, 2004). Emotions can be conveyed through various nonverbal channels, however, such as the face or voice, and little research exists about the relative ease with which people interpret this information. Eyes that grow wide with fear, a tensely uttered word, or a high-pitched scream may all signal distress, but which of these signals lead to the most accurate decoding of another's feelings? In particular, relatively little research exists about the decodability of nonlinguistic "affect vocalizations" (Scherer, 1988, 1994), such as laughter, screams, and cries. In this study, we examined observers' absolute and relative accuracy in interpreting nonlinguistic affect vocalizations as discrete emotions.

Various nonverbal emotion cues help to regulate social life at many levels, and thus can serve a multitude of overlapping social functions (Fischer & Manstead, 2008; Scherer, 1980, 1988, 1994). For example, in face-to-face conversations, many behaviors function semantically to amplify, modify, or contradict speech. Nonverbal displays also have pragmatic value, because receivers interpret many cues as indicative of emotions. Simultaneously, these behaviors also serve as appeals for receivers to behave in certain ways. Expressions of emotions such as anger or contempt, for example, communicate a sender's negative feelings toward others, while at the same time serving to increase others' social distance from the sender and reduce further interactions (see Fischer & Roseman, 2007). The extent to which nonverbal cues fulfill these various functions likely depends on receivers' accurate decoding of these cues. As such, the relative decodability of emotion signals from various nonverbal channels is a critical issue.[1]

To date, most research on emotion decoding has focused upon facial displays and speech-embedded vocal prosody. The latter term refers to suprasegmental cues, such as fundamental frequency (pitch), amplitude (volume), and rate of speech, as well as vocal timbre and tone, which can communicate affect while speaking (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2001, 2003). People can accurately infer discrete emotions from both facial cues and speech-embedded prosody at levels well above chance (e.g., Banse & Scherer, 1996; Biehl et al., 1997; Haidt & Keltner, 1999; Juslin & Laukka, 2001, 2003; Rosenberg & Ekman, 1995; Van Bezooijen, Otto, & Heenan, 1983). There also exists a long history of theoretical interest in

Skyler T. Hawk, Gerben A. van Kleef, Agneta H. Fischer, and Job van der Schalk, Department of Social Psychology, University of Amsterdam, the Netherlands.

Correspondence concerning this article should be addressed Skyler T. Hawk, Department of Social Psychology, University of Amsterdam, Roeterstraat 15, 1018 WB Amsterdam, the Netherlands. E-mail: s.t.hawk@uva.nl

[1] While affect vocalizations and speech-embedded prosody both utilize the same expressive (vocal) apparatus, we use the term *channel* throughout the article as an easy way to refer to these two different types of vocal signals.

nonlinguistic affect vocalizations (see Scherer, 1994, for an extensive review), but they have received scant attention in prior decoding research. We address this empirical gap in the present study by examining the absolute decodability of these nonlinguistic sounds, and also their decodability relative to nonverbal signaling through facial and speech-embedded prosodic cues.

Prior Literature on Affect Vocalization Decoding

Only a few empirical studies have examined emotion decoding from nonlinguistic affect vocalizations (abbreviated as *affect vocalizations*), which for the present research, we define as "short, emotional nonspeech expressions, comprising both clear nonspeech sounds (e.g., laughter) and interjections with a phonemic structure (e.g., 'Wow!')" (Schröder, 2003, p. 103). This includes both unconventionalized, "raw" sounds that arise from changes in the autonomic nervous system, occur across cultures, and show high individual variability in sound patterning (e.g., the gagging sounds accompanying contact with disgusting stimuli), as well as "emblems" that have an invariant phonemic structure, are moderated by social norms, and differ between cultures (e.g., "Yuck!"; Scherer, 1994). This term does, however, exclude linguistic interjections that comprise actual speech (e.g., "Gross!"). Although this distinction seems theoretically clear, it becomes less evident when considering actual sounds; most affect vocalizations fall somewhere in the middle of the raw-emblematic continuum, and we may expect "all sorts of mixtures" to exist (Schröder, 2003, p. 100). While some might think that this should invariably lead emblematic sounds to be more decodable than "raw" vocalizations within a given culture, prior research actually suggests that listeners can also decode many "raw" expressions with high accuracy.

In one early study, Hatfield, Hsee, Costello, Weisman, and Denney (1995, Study 2) pretested a set of 10- to 12-s sound clips in which a female encoder produced joyful laughter, sobs of sadness, throaty growls of anger, and "short, sharp cries and gasps" (p. 305) representing fear, all of which can be classified as "raw" vocalizations (see Schröder, 2003). Hatfield and colleagues obtained decoding scores for all emotions substantially higher (above 72%) than the typical 50% to 60% accuracy rate for speech-embedded prosody (see Banse & Scherer, 1996, and Scherer, 1999). Of particular note were higher scores than those usually attained for speech-based prosodic expressions of joy and fear (see Scherer, 1999). Schröder's (2003) study of affect vocalizations also showed similar scores, reporting a mean accuracy of 81% in a listening test, and found comparable results for additional emotions such as disgust and contempt. Schröder was also the first to make direct comparisons of "raw" and emblematic vocalizations. He reported instances of high decoding accuracy from both types of stimuli, meaning that even "raw" vocalizations can be highly decodable for listeners. Considering these promising results, it seems fruitful to obtain further insight into the emotions interpreted from affect vocalizations.

Research on emotion decoding has mostly focused on the discrete emotions of anger, contempt, disgust, fear, joy, sadness, and surprise, in part because of their assumed universality (e.g., Biehl et al., 1997; Elfenbein & Ambady, 2002; Keltner & Ekman, 2000; Rosenberg & Ekman, 1995, but see the critique by Russell, 1994). In the sole experimental study of affect vocalization decoding (Schröder, 2003), however, only a few of these categories were included (i.e., anger, disgust, and contempt), and others were examined only indirectly (e.g., "startle" instead of surprise) or ignored altogether (e.g., sadness). There are also facial and vocal cues for emotions such as embarrassment and pride (e.g., Banse & Scherer, 1996; Edelmann et al., 1989; Keltner, 1995; Tracy & Robins, 2004), rarely included in larger decoding studies and never before in studies of affect vocalizations. We therefore examined affect vocalizations representing embarrassment and pride, as well as the seven emotions most commonly studied in decoding research.

Nonlinguistic Affect Vocalizations Compared With Other Nonverbal Signals

Beyond the *absolute* decoding of affect vocalizations, an additional question pertains to their *relative* decodability when compared with other forms of nonverbal signaling. Evidence suggests a fairly high level of cue redundancy between emotions when expressed *within* certain channels. For example, facial displays may incorporate some of the same muscle groups in the expression of different emotions (e.g., Rosenberg & Ekman, 1995), and encoders may also use acoustic cues such as pitch or volume in similar ways to express different emotions vocally (Banse & Scherer, 1996; Juslin & Laukka, 2001, 2003). Evidence also shows, however, that there is a rather low amount of information redundancy *between* channels (e.g., Burns & Beier, 1973; Elfenbein & Ambady, 2002; Scherer, 1994; Wallbott & Scherer, 1986). Information from multiple channels can thus help to maximize chances of accurately perceiving a sender's emotions. This also raises the question of whether some forms of nonverbal behavior (including affect vocalizations) might hold certain communicative advantages over others.

Nonverbal emotion signals from different channels can be compared on the basis of their *detectability* and *discriminability* (e.g., Guilford & Dawkins, 1991). Without both of these qualities, the social functions of the behavior(s) will likely be quite limited. Vocal cues have general *detectability* advantages, because they can travel omnidirectionally and over long distances. Facial signals, on the other hand, depend on a prior visual orientation toward the sender (Scherer, 1994). Vocal cues are therefore especially effective both for eliciting attention and communicating emotions in absence of visual input (Guilford & Dawkins, 1991; Johnstone & Scherer, 2000; Scherer, 1994; Seyfarth & Cheney, 2003). Thus, people may miss out on life-saving information if they do not notice another's fearful face, but if they hear a scream, they can redirect their attention to a potential threat.

Nonverbal signals across channels also differ in their *discriminability*. Accuracy scores from prior studies suggest that facial cues often provide more discriminable emotion information than do speech-embedded prosodic signals, especially for joy, contempt, disgust, fear, embarrassment, and pride (Banse & Scherer, 1996; Biehl et al., 1997; Elfenbein & Ambady, 2002; Rosenberg & Ekman, 1995; Scherer, 1999; Scherer, Banse, Wallbott, & Goldbeck, 1991; Tracy & Robins, 2004; Van Bezooijen et al., 1983). Experiments examining different combinations of video and audio input also consistently show a predominance of visual information (Burns & Beier, 1973; Wallbott & Scherer, 1986). Thus, a sender's smile can likely signal her happiness more readily than does her joyful tone of voice.

A relative lack of distinctive acoustic cues between discrete emotions may sometimes lead to poorer decoding accuracy from speech-embedded prosody (Bachorowski, 1999; Banse & Scherer, 1996; Juslin & Laukka, 2001, 2003). For example, emotions with similar theoretical profiles of physiological arousal are more often confused than are those with disparate profiles (Bachorowski, 1999; Laukka, Juslin, & Bresin, 2005; Russell, Bachorowski, & Fernández-Dols, 2003; Van Bezooijen et al., 1983). The resulting errors can be either symmetrical (e.g., Van Bezooijen et al.'s [1983] results for disgust and contempt), or asymmetrical, such as when joy is confused for surprise, or fear for sadness, more than the reverse (e.g., Banse & Scherer, 1996; Elfenbein & Ambady, 2002). Furthermore, some emotions (e.g., disgust) may be typically expressed through short vocalizations instead of lengthy speech (Banse & Scherer, 1996; Scherer, 1994). This (as yet untested) notion implies that affect vocalizations of such emotions should be decoded more accurately than speech-embedded prosodic expressions.

When decoding affect vocalizations, receivers can rely on both the prosodic configurations of the sounds, as well as on their phonemic structures (Schröder, 2003). These two components may be differentially important for sounds representing particular emotions. For example, anger vocalizations are decoded well solely on the basis of prosody, because they have acoustic configurations that are highly differentiated from those of other emotions. Vocal expressions of disgust and contempt, on the other hand, have highly similar acoustic profiles (e.g., Banse & Scherer, 1996), and are poorly decoded from vocal prosody alone. Schröder (2003) suggested that these emotions were further differentiated by the actual phonemes (the smallest unit of speech) that encoders produced for related affect vocalizations (e.g., "Ih" for disgust and "Tse" for contempt), which elevated their decoding accuracies well above the average scores reported in prior studies of vocal prosody (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2001; Scherer, 1999; Van Bezooijen et al., 1983). It thus seems that the discrete phonemic patterning of certain affect vocalizations is a "disambiguating factor" (Schröder, 2003, p. 122) for some expressions, particularly if their vocal-prosodic configurations are poorly differentiated from those of other emotions.

Overview of Hypotheses and Methodological Considerations

From the prior empirical research on nonlinguistic affect vocalizations (Hatfield et al., 1995; Schröder, 2003), we predicted that participants would accurately decode these signals at levels significantly above chance (*Hypothesis 1*). As a function of the emotions investigated, we also expected affect vocalizations to be generally decoded more accurately than speech-embedded prosody (*Hypothesis 2*), but equal to facial expressions (*Hypotheses* 3). We also expected speech-embedded stimuli to show several decoding errors reported in prior research (*Hypothesis 4*), especially among emotions that share many overlapping acoustic features. In particular, we expected confusions between disgust and contempt, sadness ands fear, and joy and surprise. We also predicted that emotions showing these (and other) confusions in the prosodic speech stimuli would be more successfully differentiated from nonlinguistic affect vocalizations (*Hypothesis 5*).

Decoding accuracy depends on many factors, including the number of emotions included, response format, the training and number of encoders, and the amount of preselection performed prior to conducting decoding studies (Elfenbein & Ambady, 2002; Johnstone & Scherer, 2000; Russell et al., 2003). These issues influenced our choice of methods for the present research.

We chose a fixed-choice response format, in order to compare the results with the majority of prior literature. We recruited several encoders, all with professional acting experience, to avoid artifacts attributable to training, baseline vocal qualities, or facial features (Banse & Scherer, 1996; Juslin & Laukka, 2003; Scherer et al., 1991; Wallbott & Scherer, 1986). We generated three new sets of stimuli for the facial, prosodic, and affect vocalization portrayals. We produced separate stimuli for each of the channels, instead of using multichannel portrayals with different video-audio combinations produced through editing (cf. Wallbott & Scherer, 1986). We chose this method to account for the fact that it is possible to communicate affective information through single channels (e.g., on the telephone or through the face when one is not speaking), and also to ensure that actors devoted their full energy to communicating the emotion within that specific channel. We also used videos of facial displays, instead of static photos, reasoning that these were more comparable to the stream of information conveyed by vocal expressions.

We created posed stimuli, as this is overwhelmingly the norm in decoding research (see reviews by Elfenbein & Ambady, 2002 and Juslin & Laukka, 2003). Exemplars were preselected from a larger set, but all encoders were still represented for every emotion and for every channel. While using only the best-decoded exemplars can allow researchers to identify the specific facial or acoustic cues upon which decoders base their judgments (e.g., Banse & Scherer, 1996), this method also has drawbacks that would be problematic for our research aims. Namely, it has the side effect of boosting decoding accuracy rates (see Elfenbein & Ambady, 2002 for a discussion), possibly erasing naturally existing differences between channels, and also usually results in the use of different decoders for different emotion exemplars within a single study. It is clear that even with extensive preselection, well-trained actors can vary considerably both in their abilities to portray different emotions through different channels (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2003; Wallbott & Scherer, 1986), and in the cues they utilize (e.g., Banse & Scherer, 1996; Carroll & Russell, 1997). Such variability may be informative, given that untrained individuals expressing emotions in everyday life can also often be considered "actors" (Banse & Scherer, 1996), and can provide clues about whether individuals who are unsuccessful in conveying and/or discerning certain emotions through one channel have more success in others. Our choice to include every encoder for each emotion-channel combination allows a more in-depth examination of this issue, which we will address further in the Discussion.

Development of Stimuli

Encoders

Encoders were eight acting students (four men, four women), between 19 and 24 years of age, from two selective acting academies in the Netherlands.[2] All had 1 to 3 years of professional training, and had worked professionally in theater, film, and/or modeling. They were paid for their work and received instructions at least 24 hr in advance.

Production of Stimuli

Encoders were told that their goal was to produce distinct expressions for each target emotion (anger, contempt, disgust, embarrass-

[2] Admission to these acting academies is highly competitive. Last year, the academy from which most encoders were recruited received 800 applications but admitted only the top 3%. These students regularly pursue professional opportunities during training.

ment, fear, joy, pride, sadness, surprise, and a "neutral" expression) that were intense yet natural. We took care to reduce the likelihood that encoders would differentially produce low- or high-intensity versions of each emotion (see Banse & Scherer, 1996, and Juslin & Laukka, 2001, for a more detailed discussion), by discussing with encoders the meanings of the emotion labels and possible situations that could elicit such feelings. Encoders were also encouraged to recall the last time they strongly felt each emotion. They were asked to keep the same situations in mind when portraying each emotion through the distinct channels. We produced all stimuli using a digital video camera and a high-quality microphone. Encoders were filmed from the neck up while seated, with the camera directly in front of them and the microphone attached below the collarbone.

First, encoders expressed the emotions vocally but without verbal content (nonlinguistic affect vocalizations). Many actors produced a combination of sounds. For example, several exemplars for fear consisted of whimpering, audible inhalations, and screaming (see Appendix 2). Encoders also varied in the timing, pauses, length of utterance, and/or the combinations of these sounds. Second, they produced speech-embedded expressions of each emotion (termed the *speech stimuli* from hereon), saying only the unisex names "Robin and Sasha" (see Juslin & Laukka, 2003, for a review of the various standard utterances used in earlier studies). For these stimuli, we instructed encoders to avoid inserting additional sounds (such as laughs or cries) into their speech. Finally, encoders produced facial expressions for each emotion, beginning with a neutral face. Encoders portrayed each emotion twice for each channel.[3] Between the recordings, encoders received some limited feedback on the intensity and realism of their portrayals.

Preselection of Stimuli

All stimuli were edited to a length of no more than 5 s. Two raters evaluated the recordings and selected encoder portrayals from each emotion-channel combination for inclusion in the decoding study. Criteria for selection, in ascending order, were recording quality (i.e., encoders out of frame, or distorted audio samples), correspondence to discrete emotion prototypes (raters were given information, based on past research, about particular affect vocalizations, facial action units, and acoustic qualities that might characterize each target emotion), and intensity and genuineness of expressions. Interrater agreement was good (Cohen's κ = .79). The first author resolved disagreements. Eighty recordings (10 emotion categories $\times$ 8 encoders) were selected per channel.

Description of Stimuli

Table 1 contains descriptions and approximate phonetic transcriptions of the affect vocalizations, made by the first author. The first and fourth authors, certified in the Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 2002), collaboratively recorded the action codes for each visual stimulus. We also examined volume, pitch, and tempo of utterance for the affect vocalization and speech stimuli, using 43 individuals (who did not participate in the decoding task) who listened to either the volume-controlled affect vocalizations or speech stimuli, and completed 7-point Likert scales for each acoustic measure (0 = *very low*, 6 = *very high*). The descriptions of the facial cues can be viewed in the online supplemental materials for this article, as can the means, standard deviations, and comparisons of the acoustic properties.

Table 1
Descriptions of Encoders' Postselection Affect Vocalizations (Four Women, Four Men), Sorted in Descending Order by Total Frequency for Each Emotion

Emotion	%		
	Female	Male	Total
Anger			
Growling	100.00	100.00	100.00
Oh!	25.00	25.00	25.00
Contempt			
Huh/Heh	75.00	100.00	87.50
Tse/Tsh	100.00	75.00	87.50
Snort	25.00	50.00	37.50
Tongue click (Tsk)	25.00	0.00	12.50
Sigh	0.00	25.00	12.50
Disgust			
Ugh	100.00	75.00	87.50
Uah/Buah	50.00	75.00	62.50
Gagging	75.00	25.00	50.00
Eew	50.00	0.00	25.00
Ih	50.00	0.00	25.00
Embarrassment			
Laughter	100.00	50.00	75.00
Uh/Umm	50.00	75.00	62.50
Throat clear	50.00	50.00	50.00
Moan (Oh)	25.00	50.00	37.50
Sigh	0.00	25.00	12.50
Fear			
Sharp inhale	100.00	100.00	100.00
Scream	100.00	50.00	75.00
Whimper	75.00	75.00	75.00
Gulp		25.00	12.50
Joy			
Laughter	100.00	100.00	100.00
Sigh	50.00	25.00	37.50
Neutral			
Humming	100.00	100.00	100.00
Pride			
haHA!	50.00	75.00	62.50
Laughter	25.00	50.00	37.50
WooHoo!	50.00	25.00	37.50
Mmm	50.00	0.00	25.00
Yeah!	0.00	25.00	12.50
Sadness			
Crying	100.00	100.00	100.00
Sniffling	50.00	100.00	75.00
Surprise			
Heh?/Huh?	50.00	100.00	75.00
Sharp inhale	25.00	0.00	12.50
Whoa!	25.00	0.00	12.50

Method

Participants

Participants were students at a large Dutch university (N = 135; 54 male, 81 female) who received either class credits or €7 as compensation. Participants were retained only if Dutch was their native language (14 participants excluded). Data were analyzed for the 121 remaining participants (50 men, 71 women), who ranged from 18 to 26 years of age (M = 20.23, SD = 1.77).

[3] Four films were available for one encoder's facial portrayals.

Measures

Raw hit rates. Participants selected one of 10 labels for each stimulus (one for each emotion and a "none of these emotions" option; Frank & Stennett, 2001), presented in alphabetical order. We calculated "raw hit rates" for each emotion (Wagner, 1993), or the total percentage of times participants accurately decoded the stimuli. We repeated this procedure to ascertain cross-confusions. We used binomial tests to discern whether correct and incorrect labels were chosen above chance (cf. Rosenberg & Ekman, 1995).

Unbiased hit rates (Hu). Raw hit rates are problematic for comparing the decodability of different groups of stimuli, as judges might differentially favor the use of certain emotion labels. We calculated encoders' unbiased hit rates (*Hu*), using Wagner's (1993) formula. These values range from 0 to 1 (perfect score). For example, if a judge correctly decodes six of eight anger stimuli (a raw hit rate of 75%), but labels a total of 10 stimuli as "anger," her *Hu* score for anger is: $6^2/(8 \times 10) = .45$. *Hu* scores, calculated per decoder, are not calculable for individual stimuli. We therefore also calculated *Hu* per encoder, for each portrayal (cf. Goeleven, De Raedt, Leyman, & Verschuere, 2008), to shed light on the distribution of high and low scores across different encoders, emotions, and channels. If an encoder's facial expression of anger is correctly identified 20 times in a group of 40 decoders (50% raw hit rate), and decoders choose the label "anger" 25 times across all facial stimuli from that encoder, the encoder's *Hu* score for the facial expression of anger is: $(20)^2/(40 * 25) = .40$.

Procedure

Participants were randomly assigned to judge 80 stimuli in either the affect vocalizations ($n = 44$), speech-embedded prosody ($n = 37$), or facial expressions ($n = 40$) condition. All stimuli were presented via computer in a random order, held constant with respect to encoder and emotion. We reversed this ordering for 50% of participants (cf. Rosenberg & Ekman, 1995). Participants first received practice stimulus (an expression of surprise). Participants could play each recording as often as they wished.

Results

Raw Decoding Accuracy Scores

We used raw hit rates to examine the prediction that nonlinguistic affect vocalizations would be decoded with above-chance accuracy (*Hypothesis 1*), and to examine cross-confusions of stimuli in each channel. Raw hit rates also provide useful information about the scores of individual encoders. A summary of the raw and unbiased hit rates, per encoder, can be viewed in Table 2. We used binomial tests to ascertain whether the raw decoding proportions surpassed chance levels. The raw percentages of accurate and inaccurate emotion decoding, and the results of these binomial tests, can be viewed in Table 3.

We tested two different levels of chance agreement. An initial, liberal criterion (Criterion 1) was set at 10%, representing completely random chance. A second, more stringent criterion (Criterion 2) pertained to a set of categories with which each emotion is commonly confused, or shares acoustic/morphological characteristics. These criterion thresholds varied, depending on the emotion and channel in question.

Significant confusions between emotions, when portrayed through speech-embedded prosody, may exist at least partially because of similarities in theoretical arousal levels (Bachorowski, 1999; Russell et al., 2003). Therefore, we based Criterion 2 for both affect vocalizations and speech stimuli upon the arousal groupings (low, intermediate, and high) reported in Van Bezooijen and colleagues' (1983) research on speech-embedded prosody with Dutch encoders and decoders, with some minor modifications. For example, these authors classified fear into the "intermediate" arousal group, noting that confusions exist with both surprise (a higher-arousal emotion) and sadness (a lower-arousal emotion). For the sake of symmetry, we therefore included fear in all Criterion 2 groupings. Pride was also included in the "intermediate" Criterion 2 grouping. Thus, each arousal grouping consisted of four emotions ("low": sadness, fear, embarrassment, neutral; "intermediate": contempt, disgust, pride, fear; "high": anger, joy, surprise, fear), and so a criterion of 25% was set for all vocal stimuli.

In the facial domain, the Criterion 2 thresholds were adopted from Rosenberg and Ekman (1995). Levels for anger, contempt, and disgust were set at 33%, as the prototypical morphologies of these expressions share some characteristics. Similar morphological overlaps can occur between surprise and fear, and between joy and pride, suggesting a 50% threshold for each pair. Further, we thought it likely that embarrassment expressions would be confused for fear and sadness, but that these confusions would largely be asymmetrical (with embarrassment being confused for the other two categories more often than the reverse). Thus, Criterion 2 for embarrassment was 33%. No Criterion 2 was set for sad or neutral faces, as no clear alternative existed (cf. Rosenberg & Ekman, 1995).

Criterion 1. The raw scores in the confusion matrices (see Table 3) indicated that all emotions in all channels obtained accuracy scores significantly above the 10% criterion, $p < .001$. Several significant cross-confusions were also present. Affect vocalizations conveying pride were labeled both as joy and contempt, and embarrassment was labeled as "none of these emotions." The pride sounds were labeled as pride or joy at equivalent rates. Participants who viewed faces misidentified fear as surprise, and pride as contempt. Participants in the speech condition confused contempt for disgust or "none of these emotions," disgust as contempt, fear as sadness, joy as surprise, pride as contempt, and embarrassment as fear. Disgust recordings were labeled as disgust and contempt at equivalent rates. Thus, supporting Hypotheses 4 and 5, speech stimuli obtained the largest number of significant confusions, while affect vocalizations and facial stimuli were rarely confused.

Criterion 2. As shown in the confusion matrices (see Table 3), most emotion-channel combinations surpassed the stricter criteria (all p's $< .01$), except for affect vocalizations of pride. Only one significant confusion that existed at the Criterion 1 level was also significant at the Criterion 2 level, namely pride affect vocalizations labeled as joy. Thus, confusions at the lower chance level usually did not remain significant at the stricter thresholds.

Between-Channel Comparisons of Decoding Accuracy

The means and standard deviations of participants' *Hu* scores can be viewed in Table 4. As *Hu* scores are proportional values, we arcsine-transformed the scores prior to analysis (Wagner, 1993). The maximum score was thus 1.57, the arcsine of 1. We entered participants' scores as dependent variables in a 10 (emotion, within) $\times$ 3 (channel, between) $\times$ 2 (decoder gender, between)

Table 2
Raw Accuracy Scores (%) and Unbiased Hit Rates (Hu) per Encoder (F = Female, M = Male) for Each Channel

		Channel								
		Affect vocalization (N = 44)			Speech (N = 37)			Face (N = 40)		
Encoder	Emotion	Raw %	*Hu*	Mean *Hu*	Raw %	*Hu*	Mean *Hu*	Raw %	*Hu*	Mean *Hu*
F01	Anger	52.27	0.50	0.53	97.30	0.83	0.52	57.50	0.53	0.54
	Contempt	95.45	0.56		62.16	0.36		87.50	0.49	
	Disgust	100.00	0.96		72.97	0.44		57.50	0.47	
	Embarrass	65.91	0.66		64.86	0.50		67.50	0.63	
	Fear	97.73	0.45		62.16	0.41		55.00	0.39	
	Joy	88.64	0.60		48.65	0.40		100.00	0.83	
	Pride	20.45	0.12		62.16	0.51		60.00	0.58	
	Sadness	93.18	0.85		86.49	0.54		67.50	0.43	
	Surprise	29.55	0.27		100.00	0.93		85.00	0.47	
	Neutral	59.09	0.36		56.76	0.33		82.50	0.57	
F02	Anger	90.91	0.89	0.66	97.30	0.90	0.45	80.00	0.73	0.54
	Contempt	93.18	0.78		62.16	0.34		72.50	0.49	
	Disgust	100.00	0.98		62.16	0.42		75.00	0.58	
	Embarrass	61.36	0.59		56.76	0.30		77.50	0.73	
	Fear	95.45	0.89		56.76	0.35		70.00	0.50	
	Joy	93.18	0.47		37.84	0.28		90.00	0.42	
	Pride	4.55	0.01		21.62	0.10		75.00	0.63	
	Sadness	100.00	1.00		70.27	0.52		97.50	0.86	
	Surprise	86.36	0.64		91.89	0.87		2.50	0.00	
	Neutral	61.36	0.34		86.49	0.37		67.50	0.47	
F03	Anger	90.91	0.91	0.58	86.49	0.71	0.35	42.50	0.36	0.60
	Contempt	93.18	0.67		45.95	0.15		82.50	0.35	
	Disgust	100.00	0.96		29.73	0.20		80.00	0.73	
	Embarrass	38.64	0.23		43.24	0.30		65.00	0.63	
	Fear	100.00	0.83		51.35	0.34		90.00	0.65	
	Joy	90.91	0.81		56.76	0.57		97.50	0.98	
	Pride	9.09	0.05		29.73	0.23		32.50	0.21	
	Sadness	97.73	0.63		75.68	0.33		92.50	0.80	
	Surprise	72.73	0.60		97.30	0.48		90.00	0.66	
	Neutral	40.91	0.14		40.54	0.15		77.50	0.60	
F04	Anger	84.09	0.84	0.61	94.59	0.66	0.40	87.50	0.67	0.55
	Contempt	97.73	0.67		56.76	0.22		77.50	0.50	
	Disgust	97.73	0.93		18.92	0.05		72.50	0.40	
	Embarrass	54.55	0.44		62.16	0.37		52.50	0.42	
	Fear	97.73	0.84		89.19	0.60		17.50	0.18	
	Joy	68.18	0.34		40.54	0.38		87.50	0.88	
	Pride	4.55	0.01		13.51	0.11		85.00	0.72	
	Sadness	97.73	0.84		97.30	0.52		82.50	0.52	
	Surprise	90.91	0.79		97.30	0.92		87.50	0.61	
	Neutral	65.91	0.37		35.14	0.19		82.50	0.62	
M01	Anger	61.36	0.55	0.53	89.19	0.82	0.42	82.50	0.63	0.62
	Contempt	43.18	0.22		10.81	0.02		55.00	0.34	
	Disgust	77.27	0.67		8.11	0.05		97.50	0.75	
	Embarrass	43.18	0.36		62.16	0.40		50.00	0.43	
	Fear	95.45	0.77		67.57	0.56		95.00	0.90	
	Joy	84.09	0.68		59.46	0.36		100.00	0.78	
	Pride	56.82	0.46		40.54	0.28		52.50	0.44	
	Sadness	93.18	0.75		97.30	0.70		95.00	0.69	
	Surprise	97.73	0.71		97.30	0.75		67.50	0.68	
	Neutral	50.00	0.15		83.78	0.29		82.50	0.52	
M02	Anger	97.73	0.93	0.76	97.30	0.66	0.44	100.00	0.83	0.65
	Contempt	86.36	0.67		54.05	0.34		57.50	0.39	
	Disgust	100.00	0.96		48.65	0.27		85.00	0.74	
	Embarrass	38.64	0.35		64.86	0.65		72.50	0.73	
	Fear	100.00	1.00		10.81	0.02		72.50	0.66	
	Joy	97.73	0.75		37.84	0.28		85.00	0.69	
	Pride	77.27	0.67		78.38	0.61		77.50	0.60	
	Sadness	100.00	0.96		78.38	0.36		75.00	0.75	
	Surprise	90.91	0.87		100.00	0.69		97.50	0.70	
	Neutral	70.45	0.40		70.27	0.51		72.50	0.40	

(*table continues*)

Table 2 (*continued*)

		Channel								
		Affect vocalization ($N = 44$)			Speech ($N = 37$)			Face ($N = 40$)		
Encoder	Emotion	Raw %	*Hu*	Mean *Hu*	Raw %	*Hu*	Mean *Hu*	Raw %	*Hu*	Mean *Hu*
M03	Anger	100.00	1.00	0.74	91.89	0.64	0.36	97.50	0.88	0.70
	Contempt	95.45	0.67		43.24	0.19		42.50	0.30	
	Disgust	97.73	0.96		29.73	0.17		100.00	1.00	
	Embarrass	79.55	0.75		48.65	0.40		77.50	0.67	
	Fear	75.00	0.73		27.03	0.11		65.00	0.63	
	Joy	77.27	0.60		54.05	0.34		82.50	0.78	
	Pride	59.09	0.50		51.35	0.25		92.50	0.76	
	Sadness	97.73	0.82		59.46	0.57		87.50	0.71	
	Surprise	100.00	0.90		100.00	0.58		100.00	0.56	
	Neutral	70.45	0.47		70.27	0.32		77.50	0.69	
M04	Anger	59.09	0.53	0.60	64.86	0.39	0.43	90.00	0.72	0.66
	Contempt	93.18	0.81		56.76	0.29		60.00	0.55	
	Disgust	97.73	0.59		27.03	0.14		92.50	0.93	
	Embarrass	81.82	0.80		51.35	0.49		70.00	0.63	
	Fear	84.09	0.82		48.65	0.40		92.50	0.67	
	Joy	88.64	0.37		59.46	0.36		77.50	0.53	
	Pride	9.09	0.04		62.16	0.49		67.50	0.49	
	Sadness	97.73	0.86		94.59	0.59		92.50	0.81	
	Surprise	95.45	0.91		100.00	0.82		77.50	0.65	
	Neutral	52.27	0.29		78.38	0.37		85.00	0.59	

repeated-measures mixed analysis of variance (ANOVA).[4] We used Sidak-adjusted pairwise comparisons to interpret differences in between-channel scores within single emotions, and between-emotion scores within single channels. These comparisons and the rank orderings of scores can be seen in Table 4 (for easy interpretation we report the untransformed *Hu* scores).

Multivariate tests showed a main effect of channel, $F(2, 115) = 58.28$, $p < .001$, partial $\eta 2 = .50$.[5,6] Supporting Hypothesis 3, means for affect vocalizations and facial displays showed no differences ($p = .24$). Facial displays obtained higher scores than speech stimuli ($p < .001$). In support of Hypothesis 2, affect vocalizations were also decoded with higher accuracy than speech stimuli ($p < .001$). A significant main effect of emotion also existed, $F(9, 107) = 39.49$, $p < .001$, partial $\eta 2 = .76$. Specific contrasts can be viewed in Table 4. Male decoders ($M = .63$, $SD = .18$) and female decoders ($M = .66$, $SD = .18$) showed equivalent scores.

The interaction between emotion and channel was significant, $F(18, 214) = 31.98$, $p < .001$, partial $\eta 2 = .73$. Pairwise comparisons showed significant differences between channels for all emotion categories: anger, $F(2, 115) = 5.88$, $p = .004$; contempt, $F(2, 115) = 47.94$, $p < .001$; disgust, $F(2, 115) = 94.24$, $p < .001$; embarrassment, $F(2, 115) = 7.48$, $p = .001$; fear, $F(2, 115) = 65.73$, $p < .001$; joy, $F(2, 115) = 39.49$, $p < .001$; pride, $F(2, 115) = 38.84$, $p < .001$; sadness, $F(2, 115) = 36.02$, $p < .001$; surprise, $F(2, 115) = 16.00$, $p < .001$; and neutral, $F(2, 115) = 21.58$, $p < .001$. Supporting Hypothesis 5, many emotion categories significantly misinterpreted from speech stimuli (i.e., disgust, contempt, fear, and joy) —and thus responsible for the generally lower *Hu* scores in this channel—were more accurately discriminated from affect vocalizations. No additional interactions were significant.

Discussion

The judges decoded expressions of several emotions with differential accuracy, depending on the channel of communication. Nonlinguistic affect vocalizations and facial cues, in particular, were more useful than speech-embedded prosody for decoding discrete emotions. Thus, understanding others' feel-

[4] We originally included gender encoder as an additional factor, which did not change the results reported here in meaningful ways. While some differences existed between male and female encoders, these were not systematic or readily interpretable. For space considerations, these gender differences are omitted but are available upon request.

[5] We performed an additional mixed ANOVA including decoders' raw hit rates for each stimulus to examine encoder variance as an additional within-subjects factor. *Hu* scores, when calculated per decoder, are not calculable for single portrayals. All interaction effects involving encoder were significant (F's ranging from 8.44 to 13.55, all p's $< .001$, partial η^2's between .35 and .94). We do not consider these interactions further, however, because the reported results for the emotion-channel interactions were largely unaffected. Also, examining which particular encoders were effective in each emotion-channel combination is of little theoretical value. We also note the rather even distribution of encoders' raw and unbiased scores, as seen in Table 2, which we address further in the Discussion.

[6] No stimuli were omitted. We concluded that including encoders' full range of performances was a valid approach, because individuals typically differ in their abilities to portray various emotions in different ways. We repeated these analyses, omitting exemplars with *Hu* scores lower than .25, which completely replicated the pattern of between-channel rank orderings from the original analysis. We performed the analyses again, omitting stimuli with *Hu* scores lower than .50 (emotion-channel combinations with no exemplars over .50 received an unbiased score of 0). The general superiority of the affect vocalization and facial stimuli over the speech stimuli remained. The between-channel rank orderings were identical except for the following cases: For anger, all three channels were now equal, and facial stimuli for contempt and disgust became equal to affect vocalizations. Affect vocalizations of embarrassment were equal to both face and speech stimuli, and speech stimuli for surprise became equal to facial stimuli. Details of these further tests are available upon request.

ings hinges not only upon *which* emotions they express, but also often upon *how* they relay that information.

Absolute Decodability of Affect Vocalizations

Supporting Hypothesis 1, nonlinguistic affect vocalizations were decoded with high accuracy. The scores for disgust and contempt are particularly interesting, given their very low accuracies in studies on speech-embedded prosody (e.g., Banse & Scherer, 1996; Van Bezooijen et al., 1983). These scores are equivalent to the only other study of affect vocalizations including disgust and contempt (Schröder, 2003; 93.1% and 84.4%, respectively). Major confusions were quite uncommon when decoding affect vocalizations.

The current investigation adds to the scant knowledge available on affect vocalization decoding. Given the relative neglect of emotions such as embarrassment and pride in prior research, it is noteworthy that affect vocalizations of embarrassment were decoded above Criterion 2 levels, and that pride was the only emotion failing to surpass the Criterion 2 thresholds. Only affect vocalizations for pride and embarrassment held significant confusions with other categories. While visual cues for these emotions have been identified in prior studies (e.g., Keltner, 1995; Tracy & Robins, 2004), more research is needed on vocal cues.

Relative Decodability of Affect Vocalizations

The present investigation is novel in assessing the relative decodability of affect vocalizations, compared with other forms of nonverbal signaling. Both affect vocalizations and facial displays often obtained higher scores than speech stimuli (supporting Hypotheses 2 and 3). The speech stimuli showed eight significant confusions—nearly twice the number of the other two channels combined (supporting Hypothesis 4) – and largely paralleled errors reported in prior studies. In such situations, these emotions obtained higher scores from affect vocalizations (supporting Hypothesis 5). *Hu* scores for speech stimuli were never superior to affect vocalizations, but the two channels were equivalent for embarrassment, pride, surprise, and neutral exemplars. A general superiority of one channel over another likely depends on the emotions included, but the findings do suggest that judges decoded many emotions more poorly from speech stimuli than from affect vocalizations or faces.

One explanation for the higher decoding accuracy in the affect vocalization and facial channels is that encoders might have relied to a greater extent upon emblematic expressions (e.g., "Yuck!"; Scherer, 1980, 1994) that are culturally bound and convey a symbolic, invariant message. For the following reasons, however, we hesitate from fully explaining the cross-channel results (particularly between the two types of vocal stimuli) in such terms.

First, some affect vocalizations that were produced by all encoders and attained high accuracy scores would be difficult to categorize as true emblems. While sounds such as joyful laughter, fearful screams, cries of sadness, and perhaps also angry growls, are likely prototypical for their respective emotions, they probably still represent relatively "raw" expressions because of their occurrence across cultures and high transcription variability (see Schröder, 2003, and Scherer, 1994). We would suggest that many of the disgust sounds produced by encoders also lay toward the "raw" end of this continuum. To the extent that the higher accuracy scores are results of emblem usage, however, this may be additional evidence that emblems can play an important role in expressing emotions that are not associated with highly prototypical prosodic configurations in running speech (Banse & Scherer, 1996; Scherer, 1994). These findings support prior research suggesting that many "raw" and emblematic vocalizations can be decoded substantially better than speech-embedded prosody (e.g., compare Schröder, 2003 with Banse & Scherer, 1996 and Scherer, 1999), and counter the idea that emblems are always decoded more accurately than "raw" vocalizations.

Furthermore, the affective signals conveyed in everyday social life often lie in between the two extremes of the theoretical raw-emblematic continuum (Banse & Scherer, 1996; Scherer, 1980, 1994; Schröder, 2003), making all but the most clear-cut cases difficult to disentangle. Even reflexive movements that would appear easily classifiable, such as disgusted tongue protrusions, can also become incorporated into the symbolic lexicon of a culture and be utilized in strategic ways (Fridlund, 1994). The use of emblems and/or prototypes may also be especially likely when emotion communication is intentional (Buck, 1984), as is the case for the majority of research utilizing posed expressions, or when more rapid information exchange is required (such as in direct conversation; Scherer, 1980). While the assumption usually is that encoders model these cues upon more "natural" occurrences (see Banse & Scherer, 1996, and Juslin & Laukka, 2003), the artificiality of the portrayals further blurs the line between "raw" expressions and emblems. We would contend that both classes of expressions likely exert similar influences upon social interactions when decoders interpret them as symptomatic of certain emotions. Future research on affect vocalizations should gather enough portrayals from different encoders, perhaps across different cultures, in order to draw firmer conclusions both about what constitutes an emblem and their proportional use in the expression of different discrete emotions.

Comparisons With Prior Literature

We further examined the quality of our stimuli by comparing them to exemplars used in prior research. In many respects, our stimuli seem structurally similar to those from past studies. With regard to vocal stimuli, as can be seen in the online supplements, the patterns of judges' acoustic ratings paralleled previous predictions and findings pertaining to the vocal expression of discrete emotions (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2001; Scherer, 1986). For example, fear, joy, and surprise received higher overall pitch scores, and embarrassment the lowest. Fear, anger, and joy received higher volume ratings, while embarrassment was rated lowest. Joy, fear, surprise, and anger obtained the highest respective general ratings across channels, and embarrassment, contempt, and disgust the lowest (see Banse & Scherer, 1996; Juslin & Laukka, 2001, 2003; Scherer, 1986). Our facial stimuli also included a high proportion (often 75% or higher) of prototypical

Table 3
Accurate Decoding and Confusion Percentages (Raw Hit Rates) and Binomial Tests for Each Emotion Category, Sorted by Channel

Intended emotion	Perceived emotion (%)									
	Anger	Contempt	Disgust	Embarrass	Fear	Joy	Pride	Sadness	Surprise	None
Anger	**79.55**^{b}		7.39		6.82			.57		5.68
Contempt	.28	**87.22**^{b}	.57			.57	5.11	.28	.28	5.68
Disgust	1.42	.57	**94.32**^{b}	.57					1.14	1.99
Embarrass	.57	4.83		**57.95**^{b}	3.13	.85	.28	7.95	1.70	22.73^{a}
Fear	.57				**93.18**^{b}			4.26	.57	1.42
Joy		4.83		.57	.85	**86.08**^{b}	2.84	2.27	.57	1.99
Pride		17.33^{a}		1.14	1.14	34.94^{b*}	**30.11**^{a}	.28	3.41	11.65
Sadness				1.14	1.42			**97.16**^{b}		.28
Surprise		.28			9.38	.57	.28	1.14	**82.95**^{b}	5.40
Neutral		10.23	.28	4.55	.28	14.77	3.41	.57	7.10	**58.81**^{b}
Anger	**89.86**^{b}	1.35	1.01			.34	.68	1.35		5.41
Contempt	5.07	**48.99**^{b}	21.62^{a}	1.01	.68		2.03	1.01	.68	18.92^{a}
Disgust	9.46	27.70^{a*}	**37.16**^{b}	3.72	1.35	1.69	2.03	4.05	2.03	1.81
Embarrass	.34	1.69	.68	**56.76**^{b}	19.93^{a}		.68	11.82	.68	7.43
Fear	2.03			7.09	**51.69**^{b}	1.69		23.31^{a}	1.47	3.72
Joy	5.07	2.70	1.69	.68	2.03	**49.32**^{b}	11.15	1.35	19.26^{a}	6.76
Pride	1.35	16.89^{a}	2.03	.34	.34	13.51	**44.93**^{b}	.68		19.93^{a}
Sadness	.34	1.01	1.69	3.72	4.73			**82.43**^{b}	3.04	3.04
Surprise		.34		.68		.68			**97.97**^{b}	.34
Neutral	4.73	5.74	1.01	5.41	1.69	.34	3.38	12.50		**65.20**^{b}
Anger	**79.69**^{b}	5.00	1.56		.94			.94	6.88	5.00
Contempt	5.63	**66.88**^{b}	2.81	2.19		.63	4.69	.94	8.75	7.50
Disgust	.63	1.56	**82.50**^{b}	.63	1.88			.63	.63	3.75
Embarrass	3.13	.31	.63	**66.56**^{b}	5.31		.63	12.81	1.56	6.56
Fear	.63		7.50	.31	**69.69**^{b}			1.25	17.19^{a}	1.88
Joy	.31	.63				**90.00**^{b}	7.81	.31		.94
Pride		17.19^{a}				1.31	**67.81**^{b}		.31	4.38
Sadness	1.25	1.56	2.19	2.81	1.56			**86.25**^{b}	1.56	2.81
Surprise	.31	1.56	3.44	.31	5.00	11.88	.31	.31	**75.94**^{b}	.94
Neutral	3.75	2.81		.31	2.19	3.75	2.50	5.31	.94	**78.44**^{b}

Note. Boldface type is used for the values representing raw hit rate accuracy coefficients. Empty cells represent values of 0% (no confusion). Affect Vocalizations and Speech: 25%; Faces: Anger, Contempt, Disgust, Embarrassment (33%); Fear, Joy, Surprise, Pride (50%); Sadness and Neutral (None).
[a] Criterion 1 at level of $p < .05$. [b] Criterion 2 at level of $p < .01$.
[*] Cross-confusion that is equal to accurate decoding proportion.

actions described by Ekman and colleagues (2002) and others (e.g., Ekman & Friesen, 1986; Keltner, 1995; Tracy & Robins, 2004). These data suggest that encoders mainly relied on prototypical configurations for their portrayals, similar to the majority of stimuli used in prior decoding research. While we cannot draw conclusions about their natural occurrence, it is notable that encoders produced these patterns without prompting.

Decoding accuracy. We compared our accuracy rates with those from other investigations, reviews, and meta-analyses that have used similar emotions, response formats, and/or respondents. Decoding rates used for comparisons always pertained to Western decoders judging Western models, to reflect similar levels of in-group bias (see Elfenbein & Ambady, 2002). Unless noted, we compared the raw hit rates, as few studies have included *Hu* scores.

The speech stimuli scores for anger, disgust, joy, sadness, pride, embarrassment, and surprise were similar (less than 10% difference) or superior to many reports (cf. Banse & Scherer, 1996; Elfenbein & Ambady, 2002; Juslin & Laukka, 200l; Scherer, 1999; Van Bezooijen et al., 1983).[7] Similar to Juslin and Laukka (2003), speech stimuli for anger and sadness also showed greater accuracy than those for fear and joy. Of the two prior studies including contempt, our results were equal to one (Van Bezooijen et al., 1983; 42%) and lower than another (Banse & Scherer, 1996; 60%). Other differences included Scherer's (1999) and Van Bezooijen et al.'s (1983) higher scores for joy (57% and 73%, respectively), and Van Bezooijen et al.'s higher scores for disgust (50%, but we did replicate the confusion with contempt). Additionally, fear scores were lower than in Elfenbein and Ambady (2002; 63.7%), Juslin and Laukka (2001; 60%), Banse and Scherer (1996; 63%), and

[7] We compared our embarrassment scores with those for "shame" from Banse & Scherer (1996) and Van Bezooijen et al. (1983). We also compared our scores of surprise with those for "interest" from Banse and Scherer (1996).

Table 4
Means, SDs, Rank Orderings, and Pairwise Comparisons of Unbiased Hit Rates (Hu)

Emotion	*Hu* scores: Affect vocalizations *M*	*SD*	Speech *M*	*SD*	Faces *M*	*SD*	Total *M*	*SD*	Rank orderings (*M*): Affect vocalizations	Speech	Faces
Anger	$.77^{a,2}$	.19	$.70^{b,1}$	.14	$.69^{b,1,2,3}$	.15	$.72^{1}$	.17	1^{a}	2^{b}	3^{b}
Contempt	$.64^{a,3}$	.17	$.24^{c,3,4}$	.14	$.44^{b,5}$	.20	$.45^{3}$	.24	1^{a}	3^{b}	2^{c}
Disgust	$.88^{a,1}$	.13	$.24^{c,4}$	.15	$.70^{b,1,2}$	.20	$.62^{1}$	.31	1^{a}	3^{b}	2^{c}
Embarrassed	$.52^{b,3}$	.22	$.43^{b,2,3}$	.20	$.63^{a,1,2,3,4}$	.26	$.53^{2}$	.24	2^{a}	3^{a}	1^{b}
Fear	$.77^{a,2}$	.13	$.35^{c,3,4}$	.15	$.58^{b,3,4}$	.18	$.58^{2}$	.23	1^{a}	3^{b}	2^{c}
Joy	$.55^{b,3}$	.14	$.38^{c,2,3,4}$	.21	$.71^{a,1}$	.15	$.55^{2}$	.21	2^{a}	3^{b}	1^{c}
Pride	$.25^{b,4}$	.14	$.33^{b,3,4}$	.14	$.58^{a,2,3,4}$	.23	$.38^{3}$	.23	3^{a}	2^{a}	1^{b}
Sadness	$.84^{a,1,2}$	.12	$.51^{c,2}$	.12	$.72^{b,1}$	.21	$.70^{1}$	.21	1^{a}	3^{b}	2^{c}
Surprise	$.73^{a,2,3}$	.16	$.74^{a,1}$	.13	$.54^{b,4,5}$	.17	$.67^{1}$	.18	2^{a}	1^{a}	3^{b}
Neutral	$.33^{b,4}$	.17	$.35^{b,2,3,4}$	.19	$.61^{a,1,2,3,4}$	.28	$.43^{3}$	.25	3^{a}	2^{a}	1^{b}
Total	$.63^{a}$	.05	$.43^{b}$	.07	$.62^{a}$	.14	.56	.13	1^{a}	3^{b}	2^{a}

Note. Rank-orderings range from 1 (*most accurate*) to 3 (*least accurate*); Emotions with different alphabetical superscripts across columns denote significant differences between channels ($p \leq .05$), and are organized sequentially (with "a" referring to the highest-scoring channel). Emotions with different numerical superscripts within columns denote significant within-channel differences ($p \leq .05$), and are organized sequentially (with "1" referring to the highest-scoring group of emotions). All pairwise comparisons were performed with Sidak adjustments.

Scherer (1999; 61%), but similar to Van Bezooijen et al. (1983; 46%). For all emotions, our *Hu* scores were also equal or superior to Juslin and Laukka's (2001) stimuli. Thus, only our fear stimuli consistently attained lower scores. These recordings were often confused for sadness, which may be more likely among high-intensity variants of these emotions (Banse & Scherer, 1996).

Similar equivalences existed with prior decoding rates for facial expressions. Using the same standards as for the speech comparisons, our stimuli scored better than or equal to most past reports for anger, contempt, disgust, embarrassment, fear, joy, pride, and sadness (cf. Biehl et al., 1997; Ekman et al., 1987[8]; Elfenbein & Ambady, 2002; Goeleven et al., 2008; Haidt & Keltner, 1999; Rosenberg & Ekman, 1995; Scherer, 1999; Tracy & Robins, 2004). One inconsistency was a higher score for fear reported by Ekman et al. (1987; 85%), but this was the only prior study showing this advantage. Surprise showed lower scores, as compared to the reports of Biehl et al. (1997; 92.56%), Scherer (1999; 88%), and Tracy & Robins (2004; 86%), but our scores were equivalent to those reported by Elfenbein and Ambady (2002; 80.3%), Rosenberg and Ekman (71%), and Haidt and Keltner (1999; 80%). Also noteworthy were the similar accuracy rates for embarrassment and pride (as compared with Haidt & Keltner, 1999, and Tracy & Robins, 2004), which are not often included in larger decoding studies. Furthermore, our stimuli held an average 9% unbiased accuracy advantage over a recent validation (Goeleven et al., 2008) of the Karolinska Directed Emotional Faces database (KDEF; Lundqvist, Flykt, & Öhman, 1998), and our average *Hu* scores were equal or superior for most emotions. The only exception was our lower *Hu* scores for joy (but equivalent raw scores), because our decoders often labeled pride as joy.

Based on these data, we conclude that encoders used for our research generally performed on par with, and many times outperformed, the stimuli used in prior investigations. The observed structural similarities between our stimuli and those from prior studies also seemingly translated to highly comparable accuracy rates. These parallels further bolster our confidence in our selections and the legitimacy of the cross-channel comparisons.

Strengths and Limitations

The present research had several important methodological aspects, including the use of moving facial displays and portrayals by the same encoders for all emotions and channels. In the latter case, this method reduced the chance that cross-channel comparisons were due to positive biases created through extensive pretesting. Of course, this approach has its own set of limitations, as it is likely impossible to obtain a stimulus set in which every portrayal is of equivalent quality. However, we point to the distribution of each encoder's individual scores (shown in Table 2); Each encoder produced exemplars that at times were among the best, and at other times were among the worst, reflecting differences in encoders' abilities to portray particular emotions through different channels. We would argue that this is likely also the case for untrained individuals expressing emotions in everyday life, who can also be considered "actors" in many social situations (Banse & Scherer, 1996). Accounting for encoders' relative strengths and weaknesses is thus quite telling; individuals who had difficulty encoding or decoding in particular channels could utilize others with more success.

[8] Scores used from this study are averages of United States and German samples.

Extensive preselection is utilized variably within the decoding literature, but has obvious benefits when attempting to identify the particular cues associated with certain emotions (e.g., Banse & Scherer, 1996). While the choice to use each encoder for every emotion and channel may raise questions about the quality of some portrayals, a problem also exists if some encoders are strongly overrepresented. It may be fruitful for future studies to examine the consequences of each method, and to provide suggestions for which is preferable for particular research aims. While our findings should be replicated with additional stimuli produced through different methodologies, our research certainly demonstrates the potential for affect vocalizations to be important for emotion communication between individuals.

The results of decoding studies depend on both the sending skills of the encoder and decoding skills of the receiver. This research followed the convention of using posed expressions with a high proportion of prototypical components, which has been a critique of prior investigations (e.g., Russell, 1994; Russell et al., 2003), and the findings should be further supported with naturalistic stimuli. We also cannot comment on the exhaustiveness of our stimuli, because it is clear that numerous facial and vocal configurations can convey similar messages (e.g., Juslin & Laukka, 2003). Finally, encoders' success in the speech channel may vary with the standardized utterance used (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2003). Our research followed the majority of prior studies in using only one utterance (see Juslin & Laukka, 2003 for a discussion). Besides the use of names (cf. Leinonen, Hiltunen, Linnankoski, & Laakso, 1997; Linnankoski, Leinonen, Vihla, Laakso, & Carlson, 2005), sentences (e.g., Juslin & Laukka, 2001; Van Bezooijen et al., 1983) or nonsense utterances (e.g., Banse & Scherer, 1996) have been used as verbal material. While our standardized utterance produced accuracy rates and acoustic patterns similar to those of prior studies, the present results should be replicated with additional speech samples.

We instructed encoders to portray all emotions with high intensity, discussed the meanings of the emotion labels with them, and asked them to hold the same emotional situations in mind for each channel, in order to reduce the chance that encoders differentially produced high- and low-intensity variants of the same emotion (e.g., anxiety vs. panic). High-intensity facial and vocal expressions are more decodable, however (Banse & Scherer, 1996; Juslin & Laukka, 2001; Hess, Blairy, & Kleck, 1997), and the results may thus not generalize to lower-intensity stimuli. Additionally, encoders' use of sound combinations makes it difficult to discern whether affect vocalization scores resulted from specific cues (cf. Schröder, 2003). However, these stimuli attained raw accuracy scores highly similar to Schröder's (2003) study of individual sounds (79% vs. 81%). It is also possible that we would have attained different results had we used multichannel portrayals, edited differently to satisfy the different channel conditions (cf. Wallbott & Scherer, 1986). Doing so could further increase our ability to examine the relative decodability of different channels, but only for comparing the facial displays with either of the vocal channels; affect vocalizations and speech-embedded prosody cannot co-occur simultaneously (Scherer, 1994). Nevertheless, utilizing such methods in future research may be additionally informative.

Finally, it would be interesting to investigate which cues actually predict better decoding, but with only eight exemplars per emotion, problems of statistical power prevented such analyses. Future comparisons of speech and affect vocalization stimuli, in particular, could support Schröder's (2003) suggestion that affect vocalization decoding is sometimes more dependent upon phonemic structure than upon the configuration of prosodic cues.

Conclusions and Suggestions for Further Research

An increasingly popular assumption in the emotion literature is that nonverbal signals regulate interpersonal interactions because perceivers extract important social information from these cues (e.g., Fischer & Manstead, 2008; Fridlund, 1994; Keltner & Haidt, 1999; van Kleef et al., 2004). As such, the relative decodability of discrete emotions expressed through different channels is an important issue. Our findings imply that many affect vocalizations can communicate important emotion information in the absence of contextual information or visual cues. Expressions of a particular emotion likely serve the same social functions, regardless of the channel(s) used, but particular channels may also hold important communicative advantages in different contexts. For example, the louder, higher-pitched sounds of fear could make them particularly useful for long-distance signaling of a senders' distress (e.g., Johnstone & Scherer, 2000). Other affect vocalizations may be more useful in direct conversations, however, where some sounds (and facial expressions) can amplify, contradict, or fully supplant speech (Scherer, 1980, 1994). This may be particularly true for disgust and contempt, given their low scores in the speech channel. It is clear that the contextual factors eliciting particular affect vocalizations are topics for further investigation (Scherer, 1988, 1994), as are individual differences, such as age, that might influence decoding of speech-embedded prosody (Laukka & Juslin, 2007).

The results suggest that some nonlinguistic affect vocalizations can bridge the communicative disadvantages inherent in both speech-embedded prosody and facial displays. On one hand, they hold a general detectability advantage over facial cues, and may be especially effective in eliciting the attention of others. At the same time, some affect vocalizations have a discriminability advantage over speech-embedded prosody. While no theory currently predicts which emotions should be more decodable through particular nonverbal channels, the present research can help to develop an understanding of this important issue. The results of this study demonstrate that, when comparing nonlinguistic affect vocalizations and facial displays to speech-embedded prosody, a laugh or a smile can be "worth a thousand words."

References

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8,* 53–57.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614–636.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., et al. (1997). Matsumoto and Ekman's Japanese and Caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21,* 3–21.

Buck, R. (1984). *The communication of emotion.* New York: Guilford Press.

Burns, K. L., & Beier, E. G. (1973). Significance of vocal and visual channels in the decoding of emotional meaning. *Journal of Communication, 23,* 118–130.

Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology, 72,* 164–176.

Edelmann, R. J., Asendorpf, J., Contarello, A., Zammuner, V., Georgas, J., & Villanueva, C. (1989). Self-reported expression of embarrassment in five European cultures. *Journal of Cross Cultural Psychology, 20,* 357–371.

Ekman, E., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion, 10,* 159–168.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the Judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53,* 712–717.

Ekman, P., & Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system: The manual on CD-ROM. Instructor's Guide.* Salt Lake City: Network Information Research Co.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203–235.

Fischer, A. H., & Manstead, A. S. R. (2008). Social functions of emotion. In M. Lewis, J. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed.). New York: Guilford Press.

Fischer, A. H., & Roseman, I. J. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. *Journal of Personality and Social Psychology, 93,* 103–115.

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology, 80,* 75–85.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view.* New York: Academic Press.

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion, 22,* 1094–1118.

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behavior, 42,* 1–14.

Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition and Emotion, 13,* 225–266.

Hatfield, E., Hsee, C. K., Costello, J., Weisman, M. S., & Denney, C. (1995). The impact of vocal feedback on emotional experience and expression. *Journal of Social Behavior and Personality, 10,* 293–312.

Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior, 21,* 241–257.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland (Eds.), *Handbook of emotions* (2nd ed., pp. 220–235). New York: Guilford Press.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 4,* 381–412.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68,* 441–454.

Keltner, D., & Ekman, P. (2000). Emotion: An overview. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology* (Vol. 3, pp. 162–167). Washington, DC: APA.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13,* 505–521.

Laukka, P., & Juslin, P. N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation and Emotion, 31,* 182–191.

Laukka, P., Juslin, P. N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion, 19,* 633–653.

Leinonen, L., Hiltunen, T., Linnankoski, I., & Laakso, M. L. (1997). Expression of emotional-motivational connotations with a single-word utterance. *Journal of the Acoustic Society of America, 102,* 1853–1863.

Linnankoski, I., Leinonen, L., Vihla, M., Laakso, M. L., & Carlson, S. (2005). Conveyance of emotional connotations by a single word in English. *Speech Communication, 45,* 27–39.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF).* Stockholm: Department of Neurosciences Karolinska Hospital.

Rosenberg, E. L., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion, 19,* 111–138.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin, 115,* 102–141.

Russell, J. A., Bachorowski, J-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expression of emotion. *Annual Review of Psychology, 54,* 329–349.

Scherer, K. R. (1980). The functions of nonverbal signs in conversation. In R. N. St. Clair & H. Giles (Eds.), *The social and psychological contexts of language* (pp. 225–244). Hillsdale, NJ: Erlbaum.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143–165.

Scherer, K. R. (1988). On the symbolic functions of vocal affect expression. *Journal of Language and Social Psychology, 7,* 79–100.

Scherer, K. R. (1994). Affect bursts. In S. H. M. van Goozen, N. E. van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 161–193). Hillsdale, NJ: Erlbaum.

Scherer, K. R. (1999). Universality of emotional expression. In D. Levinson, J. J. Ponzetti, Jr., & P. F. Jorgenson (Eds.), *Encyclopedia of human emotions* (pp. 669–674). New York: Macmillan.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15,* 123–148.

Schröder, M. (2003). Experimental study of vocal affect bursts. *Speech Communication, 40,* 99–116.

Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology, 54,* 145–173.

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science, 15,* 194–197.

Van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross Cultural Psychology, 14,* 387–406.

van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology, 86,* 57–76.

Wagner, H. L. (1993). On measuring performance in category judgment studies on nonverbal behavior. *Journal of Nonverbal Behavior, 17,* 3–28.

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51,* 690–699.

Received November 16, 2007
Revision received November 26, 2008
Accepted December 11, 2008 ■